# Measurements of BB Angular Correlations based on Secondary Vertex Reconstruction at sqrt(s) = 7 TeV in CMS

Published: JHEP03(2011)136 (arXiv:1102.3194)

A.Rizzi (ETH Zurich)
on behalf of the CMS Collaboration

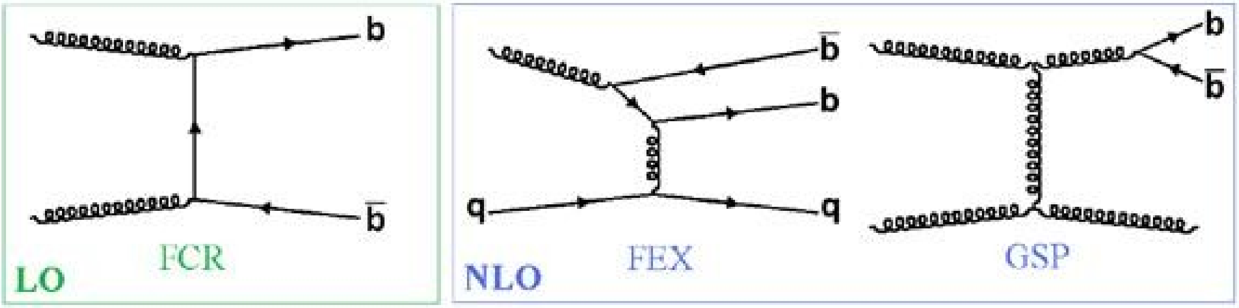
DIS2011, J.Lab ,Newport News 11-15 Apr 2011

# Outline

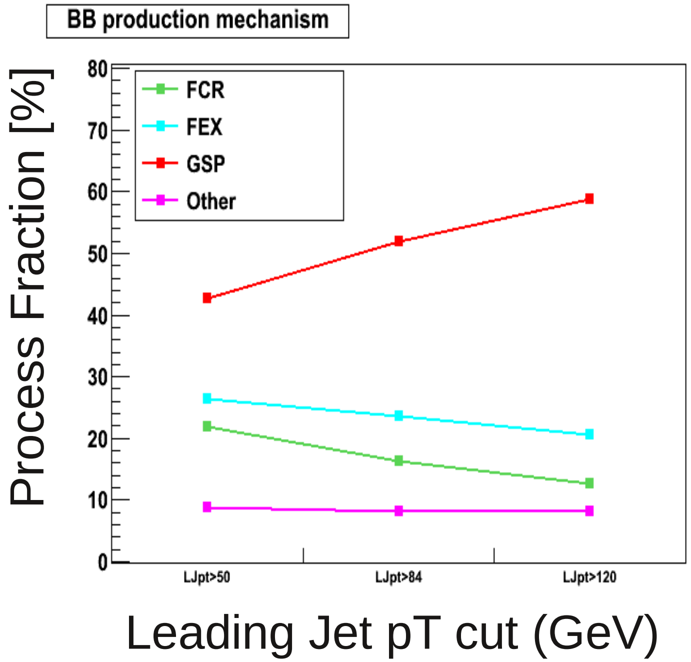
- Introduction
- Secondary Vertices
- Event selection
- Results
- Systematic uncertainties

# Introduction

- We study differential properties of B production at LHC (e.g. how much gluon splitting?):
  - **It is is a good test of perturbative QCD**
  - **It allows to verify the prediction for B production that is a background to searches**
- B-production at LHC has significant contribution from NLO diagrams
  - Flavour Creation (FCR)
  - Flavor Excitation (FEX)
  - Gluon splitting (GSP)

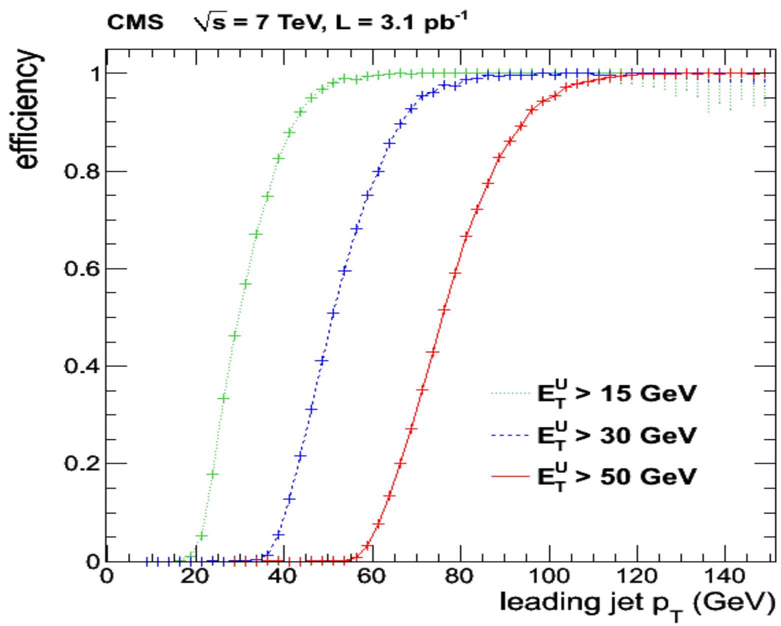


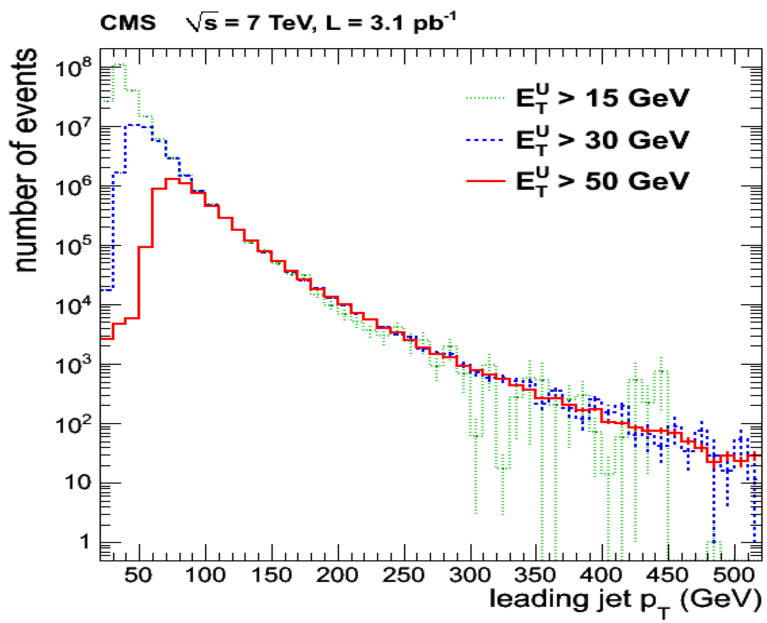

- Different processes have different properties of the produced BB pair
  - Different kinematics
  - **Different angular separation**
- In this study we measure the angular correlation between two B hadrons
- In order to highlight the different mechanism the study is repeated at three different event energy scales, determined by the leading jet in the event
  - At different energy scales the contributions from LO and NLO diagrams are different



# Trigger and leading jet bins

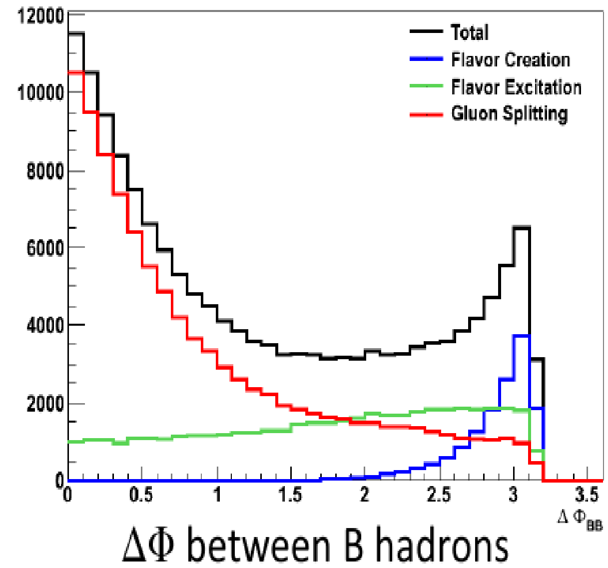
- Three bins in leading jet pT are defined and the correlation is studied for each bin
- Single jet trigger are used
  - unbiased trigger selection with respect to the offline analysis
- The pt bins are defined such that the three single jet triggers are fully efficient (99%)
  - Offline jets energy are Particle Flow jets with energy correction
  - Online jets are calorimetric with *uncorrected* energy
  - Online-Offline combinations:
    - Offline Et > 56 GeV , Online Et > 15 GeV
    - Offline Et > 84 GeV , Online Et > 30 GeV
    - Offline Et > 120 GeV, Online Et > 50 GeV

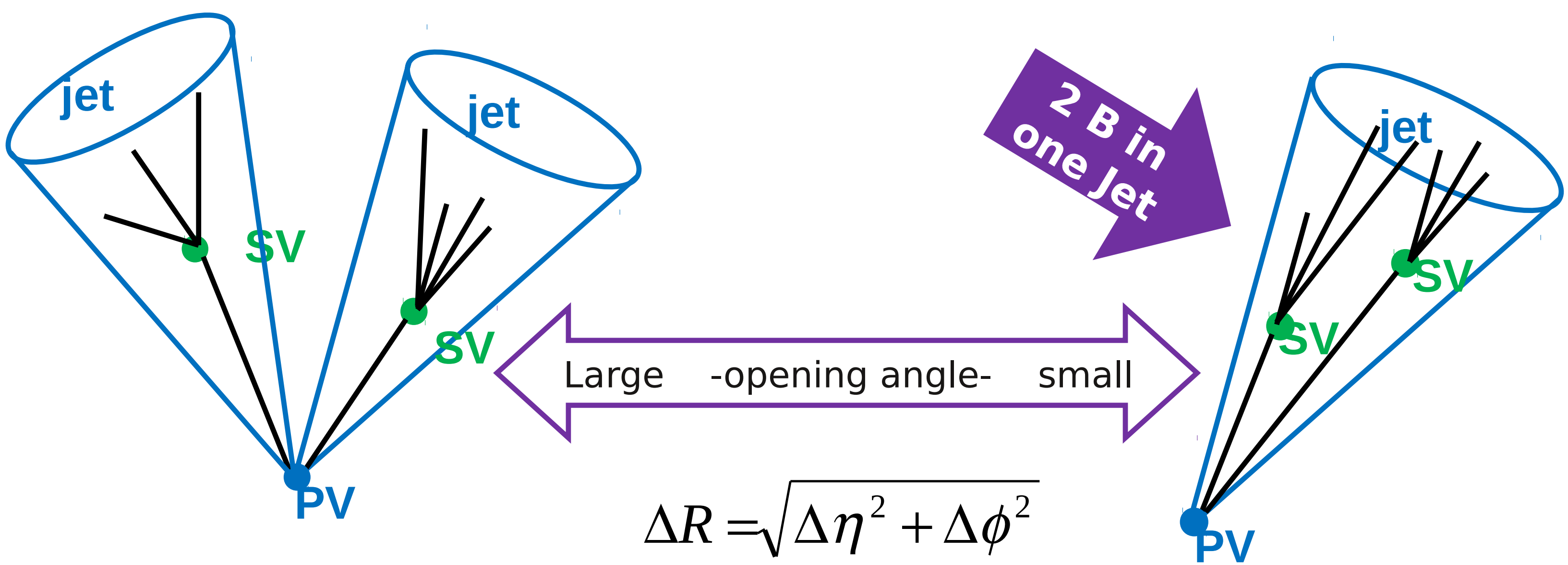


# Measurement of the angle

- While LO processes mainly contribute to the back to back topology, the GSP mechanism gives the largest contribution at small angle
- A simple approach using the jets originating from the two B as estimator of the directions would not work if the two B are in the same jet
- <u>Small angle region cannot be measured with jets</u>
  - Use Secondary Vertices and PV-SV vector as B-hadron direction estimator
  - Need to reconstruct the SV also if the 2 B hadrons are in the same jet (dedicated algorithm developed)

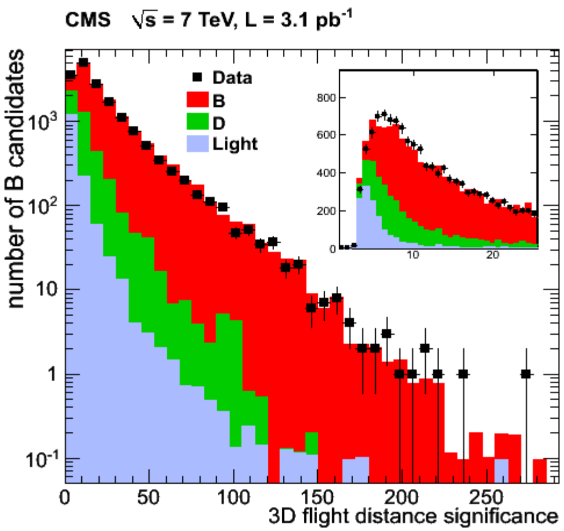


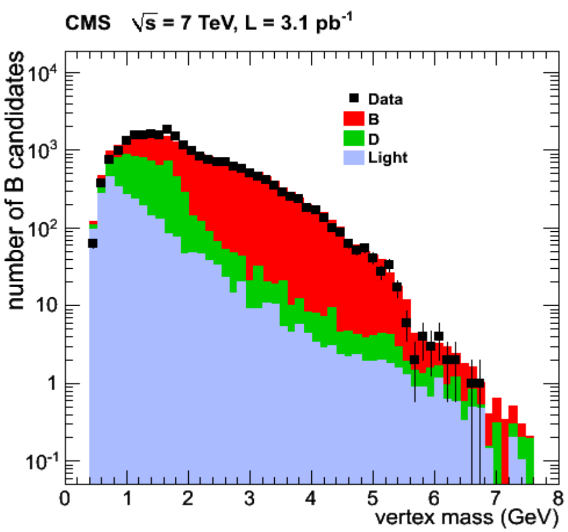



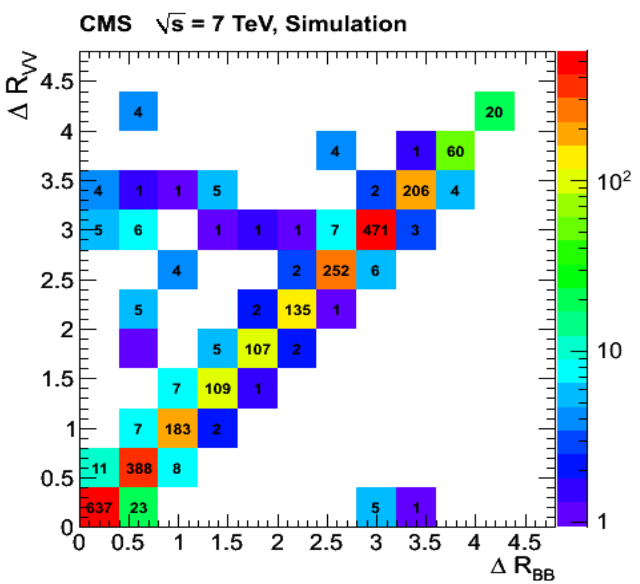

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

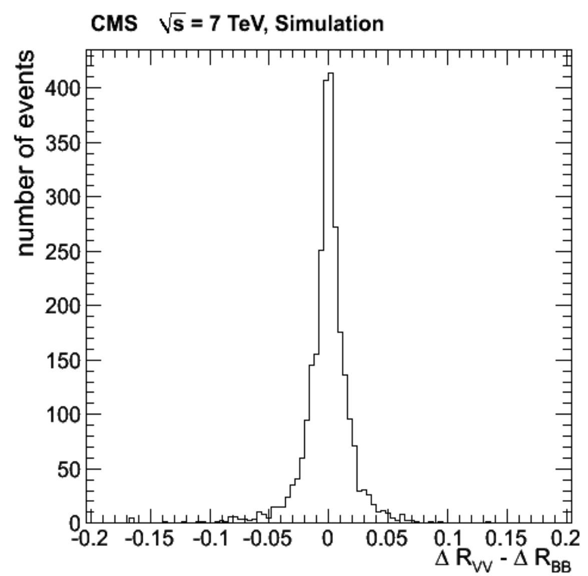
# B-hadron reconstruction

- B-Candidates are created for secondary vertices passing mass and flight distance significance cuts
  - Distributions for events with only one B checked on data
- Events with exactly two B-hadrons are used to measure the angular correlation
- Efficiency and purity are corrected from MC simulation
  - Cross checks on data are performed to estimate the systematic uncertainty arising from the corrections (see later)
- The resolution expected from MC is a factor 20 lower than the used bin width, no unfolding applied

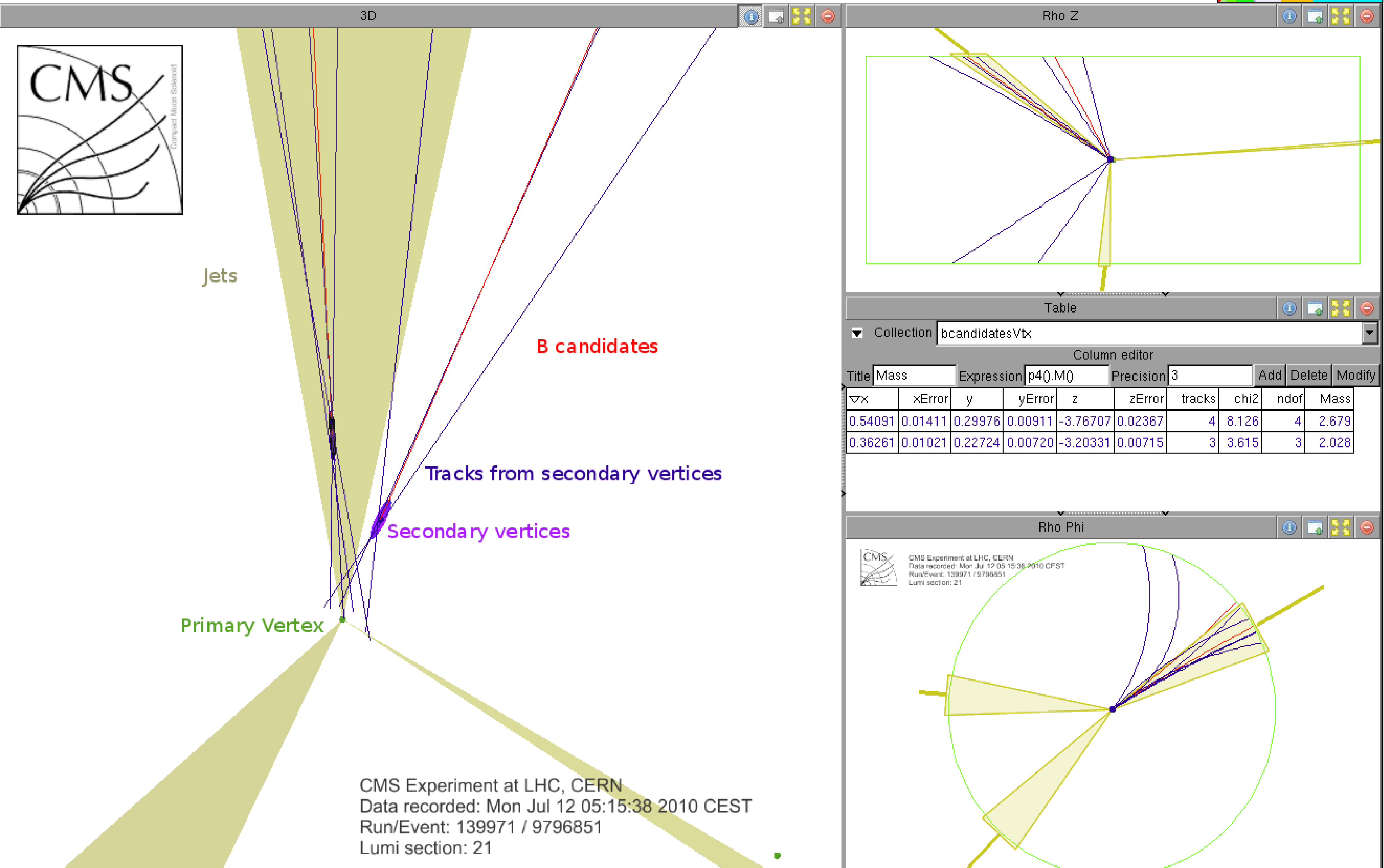


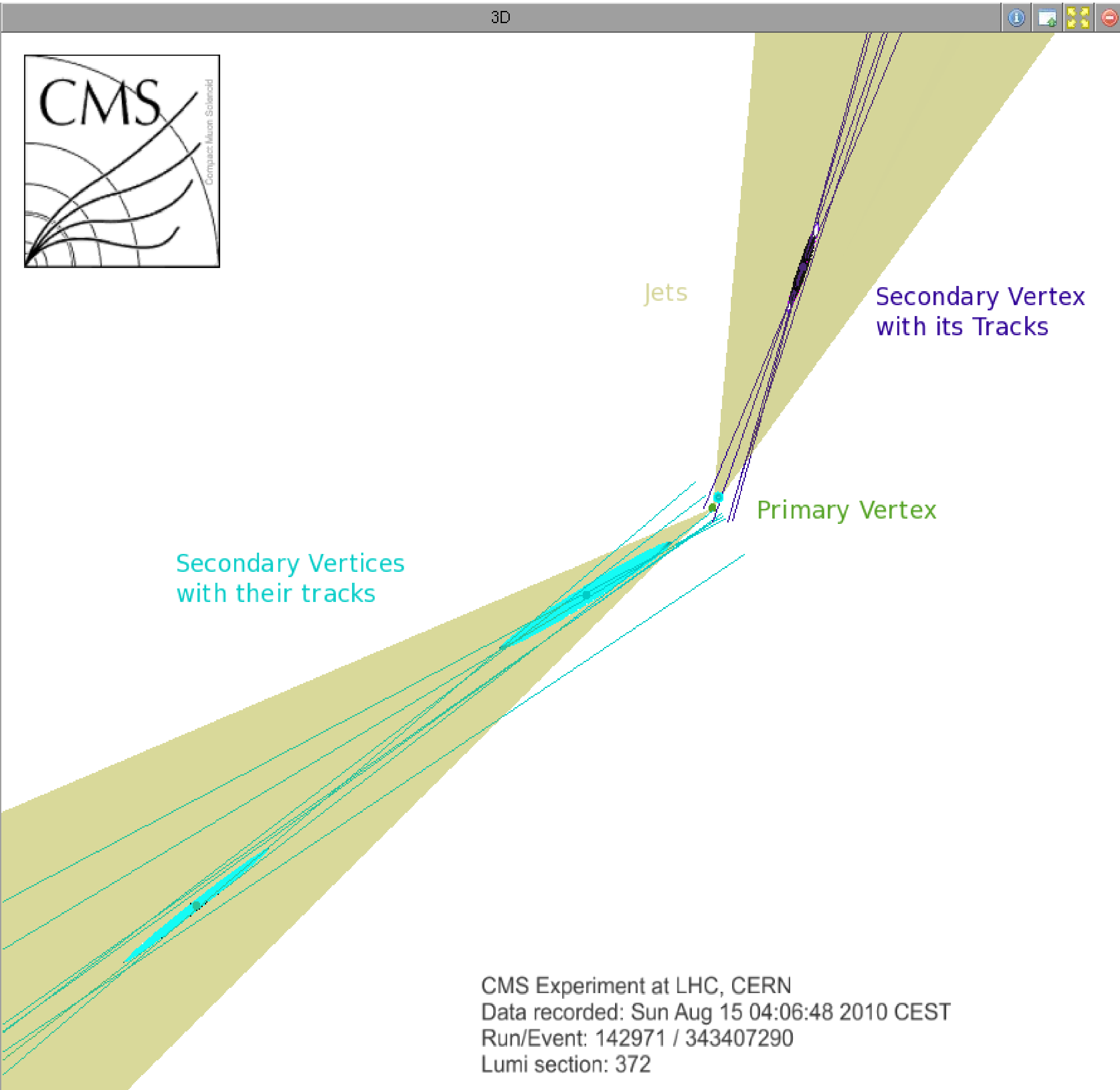



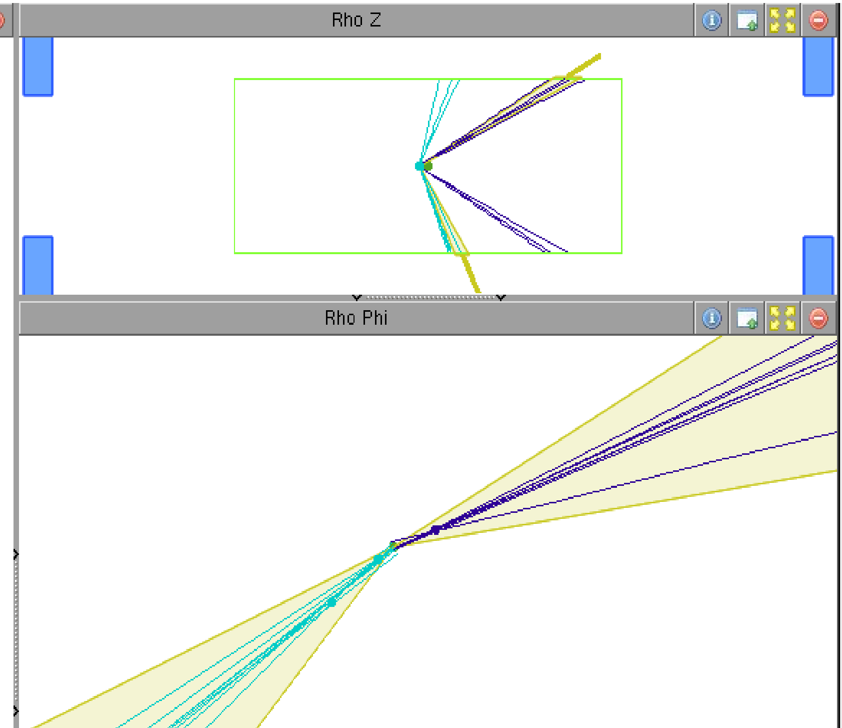
# Event display




| x | xError | y | yError | z | zError | tracks | chi2 | ndof | Mass |
|---|---|---|---|---|---|---|---|---|---|
| 0.54091 | 0.01411 | 0.29976 | 0.00911 | -3.76707 | 0.02367 | 4 | 8.126 | 4 | 2.679 |
| 0.36261 | 0.01021 | 0.22724 | 0.00720 | -3.20331 | 0.00715 | 3 | 3.615 | 3 | 2.028 |

# B->D decay




Table

Collection selectedVertices2

| x | xError | y | yError | z | zError | tracks | chi2 | ndof | Mass |
|---|---|---|---|---|---|---|---|---|---|
| -0.11300 | 0.02296 | -0.13555 | 0.01568 | -13.43550 | 0.01293 | 5 | 3.822 | 5 | 1.479 |
| -0.74276 | 0.01959 | -0.66900 | 0.01641 | -13.11158 | 0.01011 | 5 | 1.842 | 7 | 2.024 |
| 0.66114 | 0.01722 | 0.23518 | 0.00724 | -12.48443 | 0.03243 | 7 | 9.102 | 7 | 2.055 |

Table

Collection bcandidatesVtx

| x | xError | y | yError | z | zError | tracks | chi2 | ndof | Mass |
|---|---|---|---|---|---|---|---|---|---|
| 0.66114 | 0.01722 | 0.23518 | 0.00724 | -12.48443 | 0.03243 | 7 | 9.102 | 7 | 2.055 |
| -0.74276 | 0.01959 | -0.66900 | 0.01641 | -13.11158 | 0.01011 | 5 | 1.842 | 7 | 2.024 |

# Results

The BB production angular correlation is measured and reported as follow:

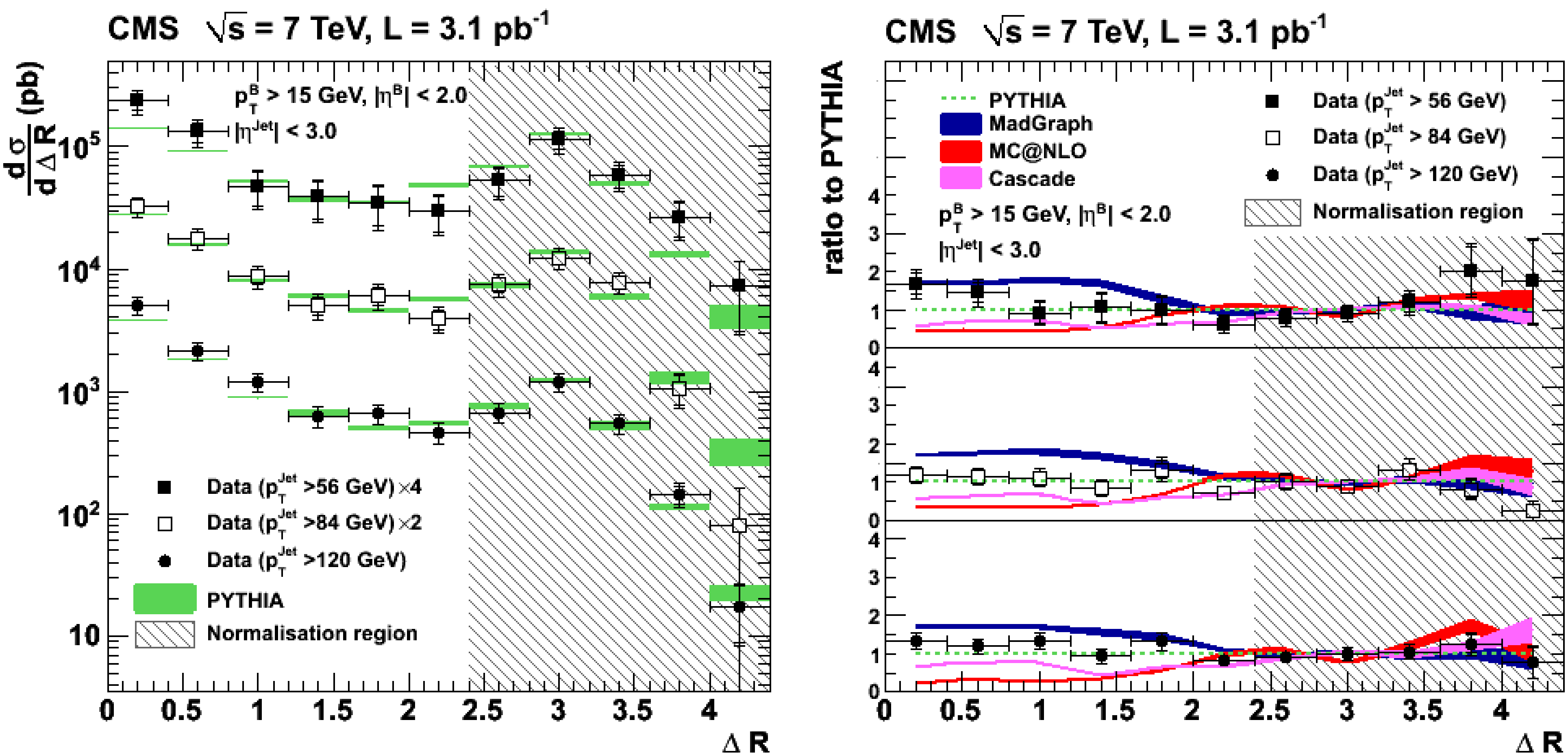
- Differential cross section distributions in $\Delta R$ and $\Delta\phi$
- In three different leading jet pt bins (event energy scale)
- Using a total luminosity of 3.1/pb (with prescales the effective lumi are 3.1/pb, 0.3/pb, 0.03/pb in the three jet energy bins)
- Two different normalization of the MC are shown:
  - **Absolute normalization** (using absolute efficiency and luminosity → large uncertainties)
  - **Relative normalization to the “back-to-back” region** ($\Delta R > 2.4$)
    - Does not depend on absolute efficiency and luminosity
    - Normalize to a well defined phase space where LO diagrams dominate
- Comparison of data with pythia prediction
- Comparison of data/pythia ratio with other-MC/pythia ratios
- Highlight of GSP dominance at high leading jet pt
  - Show $\Delta R<0.8/\Delta R>2.4$ ratio (and asymmetry) as a function of the $p_T$ of the leading jet

All results are given for the kinematic phase space where both B hadron satisfy the following conditions:

$$|\eta| < 2.0 \; , \; p_T > 15 \text{ GeV}$$

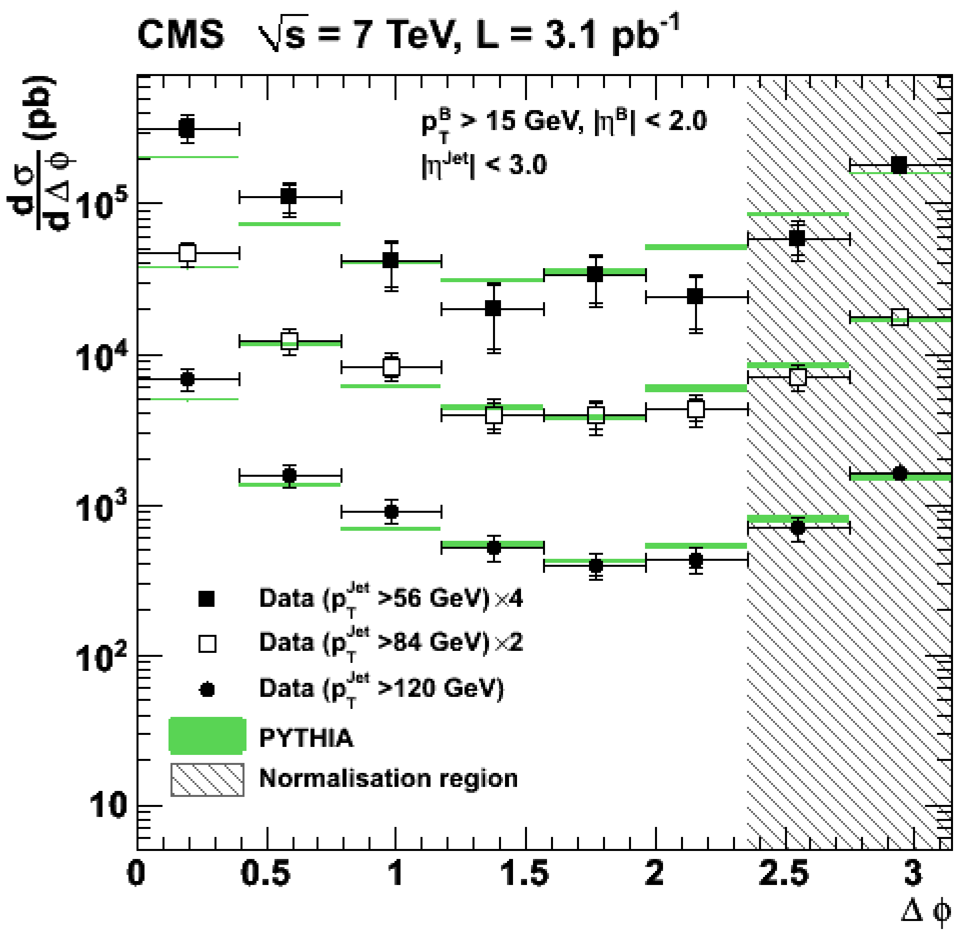
# Differential DR distribution

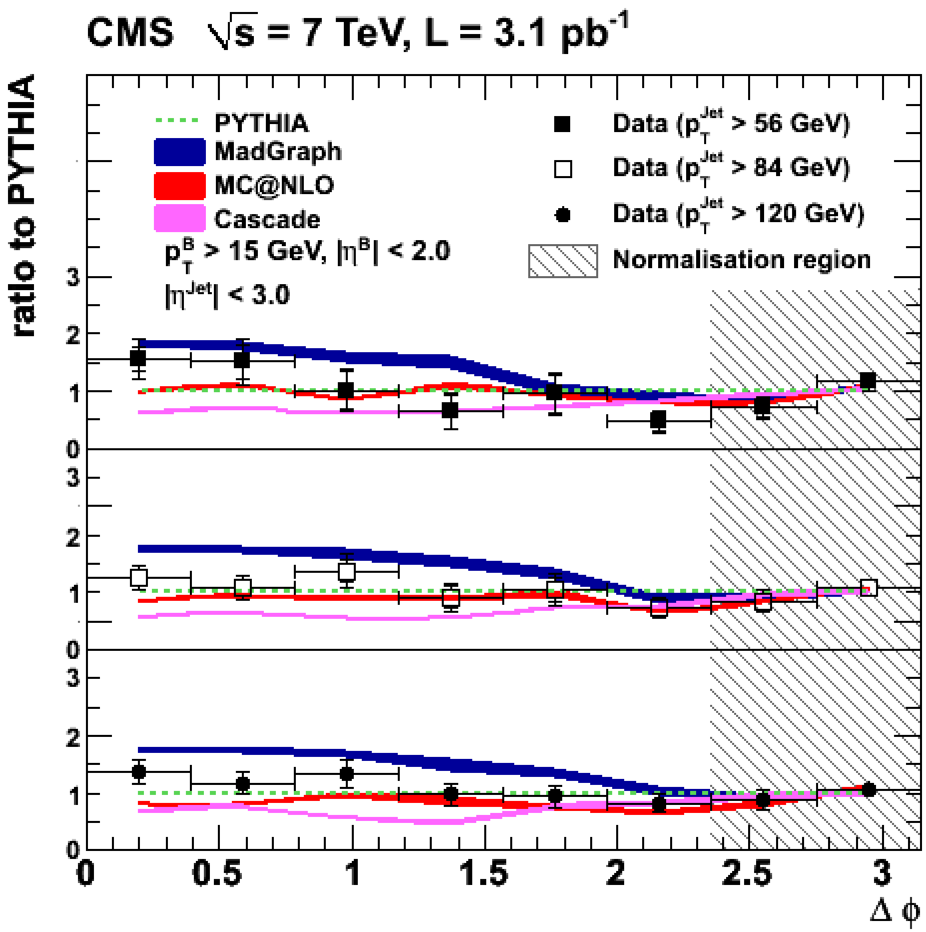
- Relative normalization to the “back to back” region
- Most generators underestimate the small angle production with respect to the back to back configuration
- Madgraph slightly overestimate it



# Differential Δϕ distribution

- Relative normalization to the “back to back” region
- Same conclusions as for ΔR

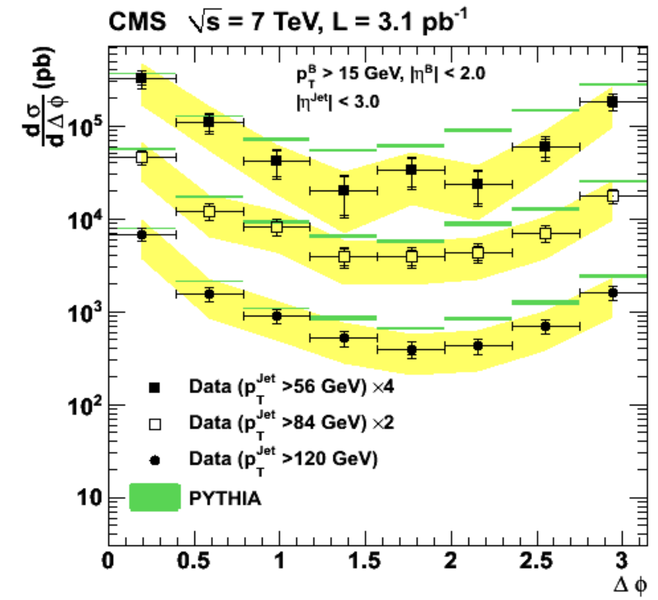


# Results

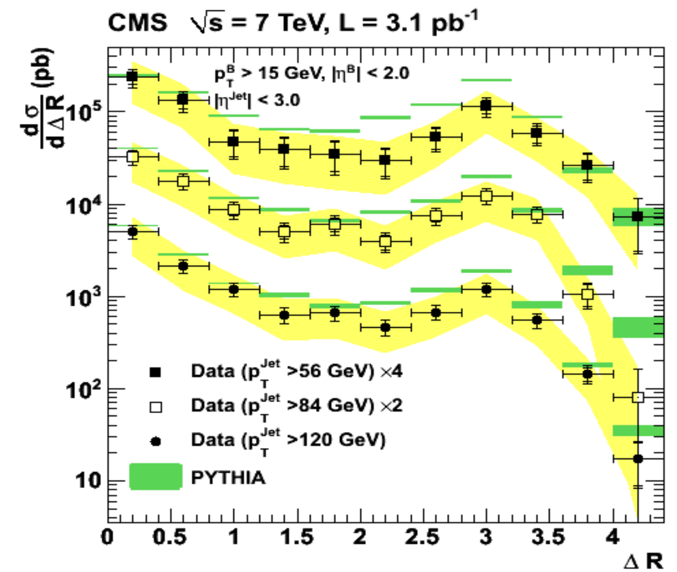
- Absolute normalization systematics is dominated by the B-Candidate reconstruction efficiency (uncertainty ~20%), requiring two B and adding luminosity uncertainty, the total is about 45% (yellow bands)

- Summary of the results:



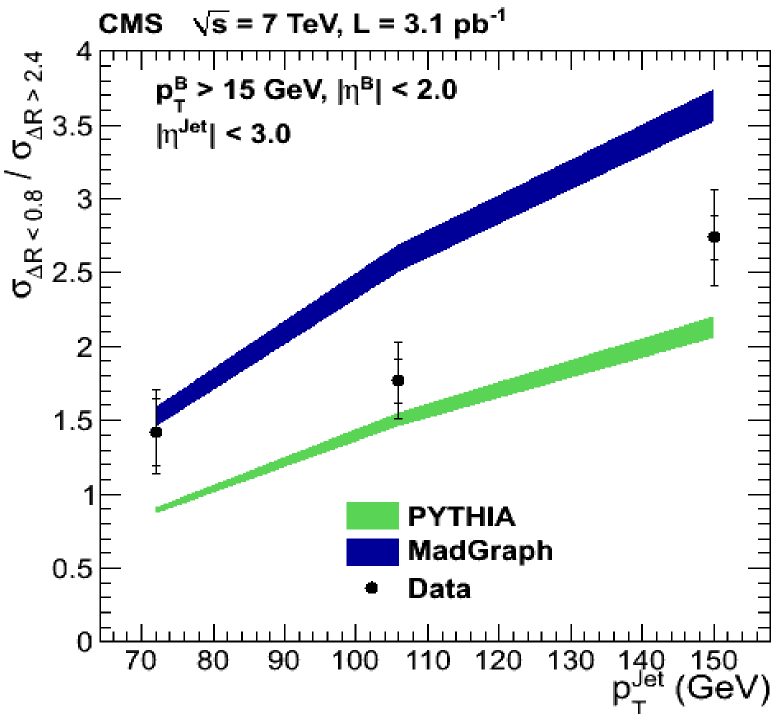

| Jet $p_T$ | | | | | | $\rho_{\Delta R} = \sigma_{\Delta R<0.8} / \sigma_{\Delta R>2.4}$ | | |
|---|---|---|---|---|---|---|---|---|
| Cut (GeV) | $\langle p_T \rangle$ (GeV) | $\sigma_{\Delta R<0.8}$ (nb) | $\sigma_{\Delta R>2.4}$ (nb) | $\langle \epsilon \rangle$ (%) | $\langle P \rangle$ (%) | Data (stat+sys) | PYTHIA (stat) | MADGRAPH (stat) |
| $> 56$ | 72 | $37 \pm 18$ | $26 \pm 12$ | 7.4 | 84.9 | $1.42 \pm 0.29$ | $0.89 \pm 0.02$ | $1.53 \pm 0.07$ |
| $> 84$ | 106 | $10.0 \pm 4.8$ | $5.7 \pm 2.7$ | 9.3 | 84.6 | $1.77 \pm 0.26$ | $1.51 \pm 0.05$ | $2.60 \pm 0.09$ |
| $> 120$ | 150 | $2.9 \pm 1.4$ | $1.0 \pm 0.5$ | 10.7 | 83.2 | $2.74 \pm 0.32$ | $2.13 \pm 0.07$ | $3.64 \pm 0.11$ |

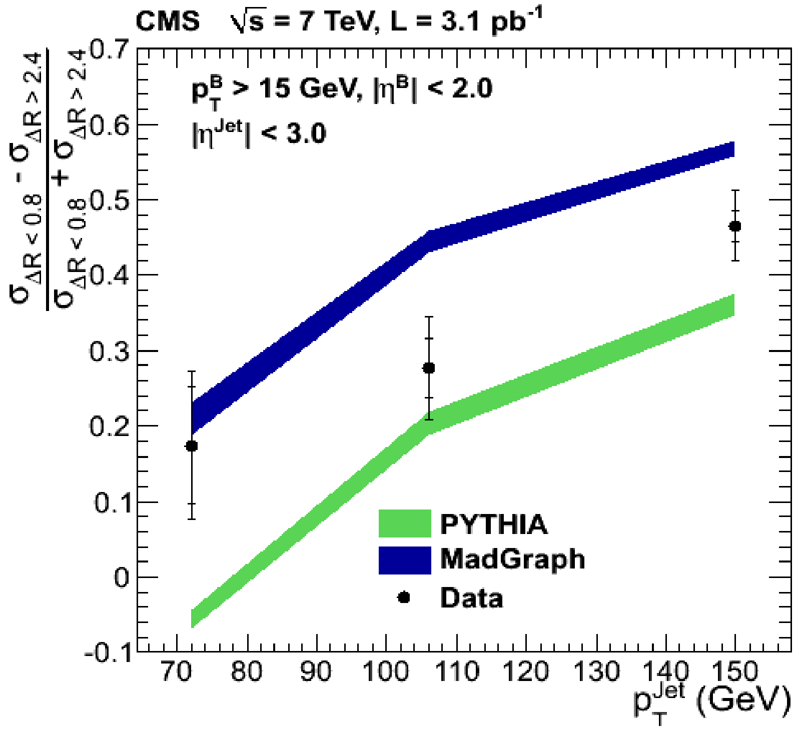
| Jet $p_T$ | | | | | | $\rho_{\Delta\phi} = \sigma_{\Delta\phi<\frac{1}{4}\pi} / \sigma_{\Delta\phi>\frac{3}{4}\pi}$ | | |
|---|---|---|---|---|---|---|---|---|
| Cut (GeV) | $\langle p_T \rangle$ (GeV) | $\sigma_{\Delta\phi<\frac{1}{4}\pi}$ (nb) | $\sigma_{\Delta\phi>\frac{3}{4}\pi}$ (nb) | $\langle \epsilon \rangle$ (%) | $\langle P \rangle$ (%) | Data (stat+sys) | PYTHIA (stat) | MADGRAPH (stat) |
| $> 56$ | 72 | $42 \pm 20$ | $24 \pm 12$ | 7.4 | 84.9 | $1.78 \pm 0.36$ | $1.15 \pm 0.15$ | $2.07 \pm 0.10$ |
| $> 84$ | 106 | $11.5 \pm 5.5$ | $4.9 \pm 2.3$ | 9.3 | 84.6 | $2.37 \pm 0.36$ | $1.95 \pm 0.25$ | $3.41 \pm 0.12$ |
| $> 120$ | 150 | $3.3 \pm 1.6$ | $0.9 \pm 0.4$ | 10.7 | 83.2 | $3.64 \pm 0.46$ | $2.73 \pm 0.32$ | $4.79 \pm 0.15$ |



# Results

- FCR/GSP ratio as a function of the event energy scale:
  - ΔR > 2.4 (FCR region)
  - ΔR < 0.8 (GSP region)




- **Pythias underestimates the GSP/FCR, while MadGraph slightly overestimates it**
- **Both generators reproduce quite correctly the leading jet pT trend**

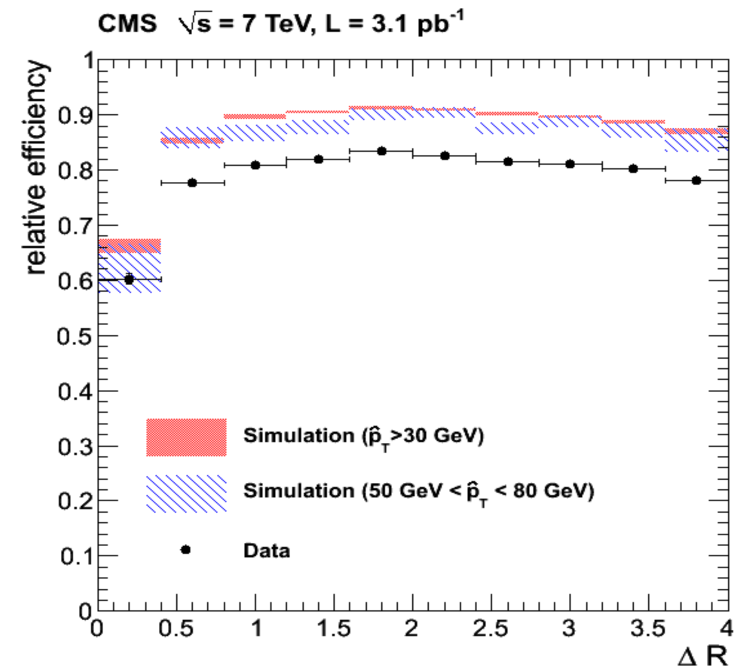
# Systematic uncertainties

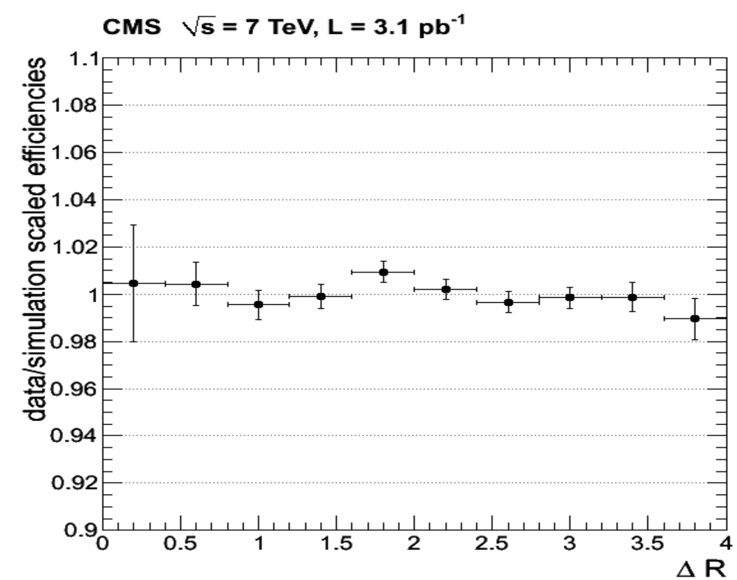
- Two types of systematic uncertainties:
  - Uncertainties affecting only the total cross section
    - Not relevant for the angular distribution, large uncertainties from average efficiency correction
  - Uncertainties affecting the shape of the angular distributions
    - Quantify this uncertainties as the variation in the ratio between GSP and FCR regions ($\Delta R < 0.8$ and $\Delta R > 2.4$)
    - Additional bin-by-bin uncertainty from limited MC statistics
  - **Data-driven cross checks:**
    - Relative efficiency (mixing 2 events with 1B each)
    - MC prediction of the pt spectrum

| Source of uncertainty in shape | Change in $\rho_{\Delta R} = \sigma_{\Delta R<0.8}/\sigma_{\Delta R>2.4}$ (%) | | |
|---|---|---|---|
| | Leading jet $p_T$ bin (GeV) | | |
| | $> 56$ | $> 84$ | $> 120$ |
| Algorithmic effects (data mixing) | 2.0 | 2.0 | 2.0 |
| B hadron kinematics ($p_T$ of softer B) | 8.0 | 7.0 | 4.0 |
| Jet energy scale | 6.0 | 6.0 | 6.0 |
| Phase space correction | 2.8 | 2.8 | 2.8 |
| Bin migration from resolution | 0.6 | 1.3 | 2.1 |
| Subtotal shape uncertainty | 10.6 | 9.9 | 8.3 |
| MC statistical uncertainty | 13.0 | 13.0 | 13.0 |
| Total shape uncertainty | 16.8 | 16.4 | 15.4 |

# Algorithmic efficiency loss at small ΔR (Event mixing)

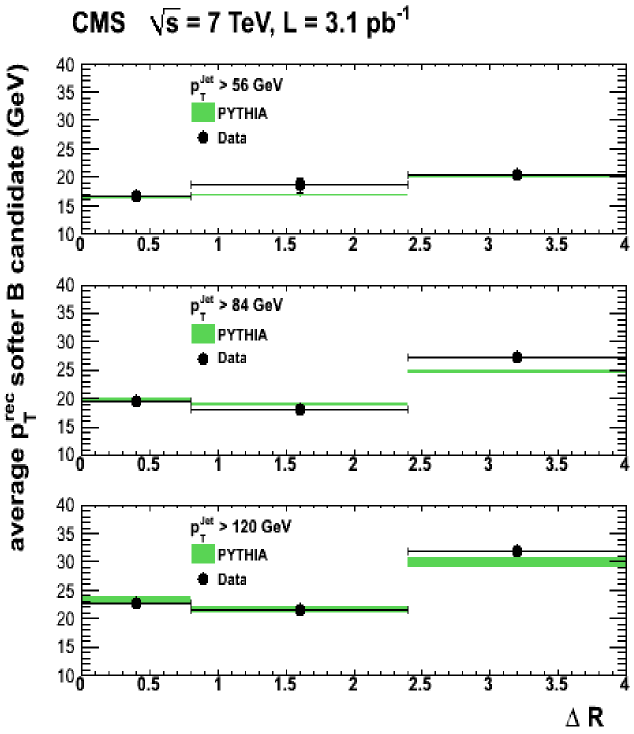
- In order to verify that the small ΔR efficiency loss is well modeled in MC we used both on data and MC an event mixing technique
- Events are pre-selected if they contain at least one B-candidate
- Pairs of events are *mixed* at the level of the electronics readings if their Primary Vertices are within the typical PV resolution (20um)
- The mixed event is re-reconstructed, re-running tracking and secondary vertex reconstruction
- A *relative efficiency* is defined by counting the fraction of mixed events where the two B candidates from the two original events are re-reconstructed
- The shape of the MC and Data *relative efficiencies* is compared and used to set a systematic uncertainty (2%)

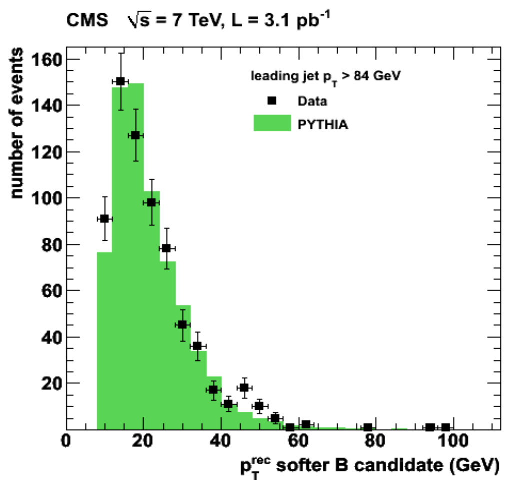


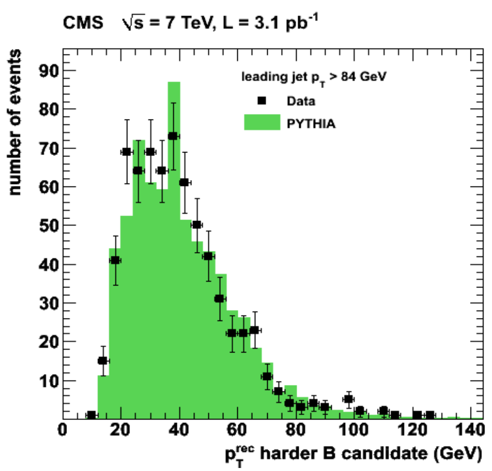



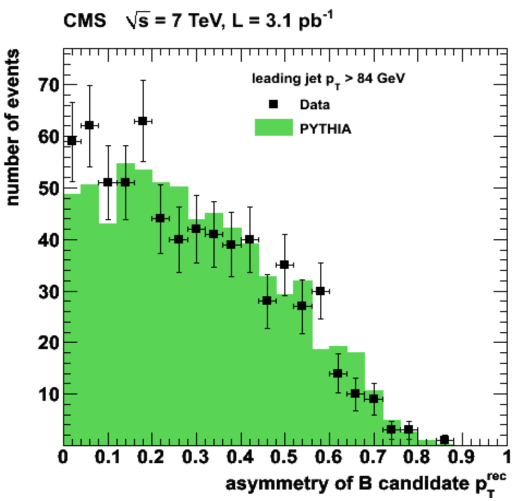
# B-hadrons kinematics

- The MC derived efficiency correction can be wrong if the spectrum of B-hadrons, for a given ΔR is not well simulated
  - The efficiency has a quite large pt dependency at low momentum
- Cross checks on the momenta of the reconstructed B-candidates have been performed:
  - Distribution of the momentum asymmetry between the two B
  - Distribution of the pT of softer and harder B
  - Trend of the mean pT for the softer B as function of ΔR
- The discrepancy are convoluted with an estimate of the efficiency vs pT dependency to compute the systematics



# Conclusions

- The first LHC measurement of BB angular correlations using secondary vertices and probing the small angle region has been performed
- Predictions from pQCD can be tested and MC generator tuned to reproduce the measured differential distributions
- The production in the collinear regime is dominant at high energy scales
- Existing Monte Carlo such as Pythia and MadGraph are found to reproduce the measured shape within 30-50% (ratio  of the two angular region)
  - Predictions are quite different for the different generators
- First steps in the understanding of one of the main background for searches with bb final states

# Backup

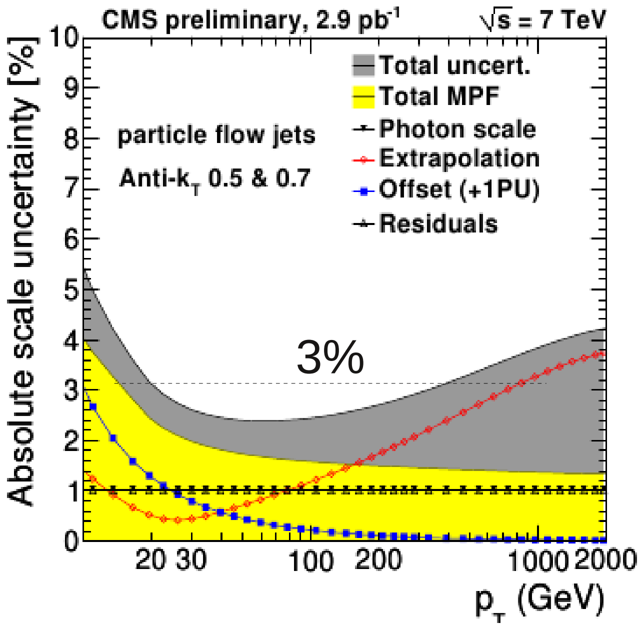
# Reconstruction phase space

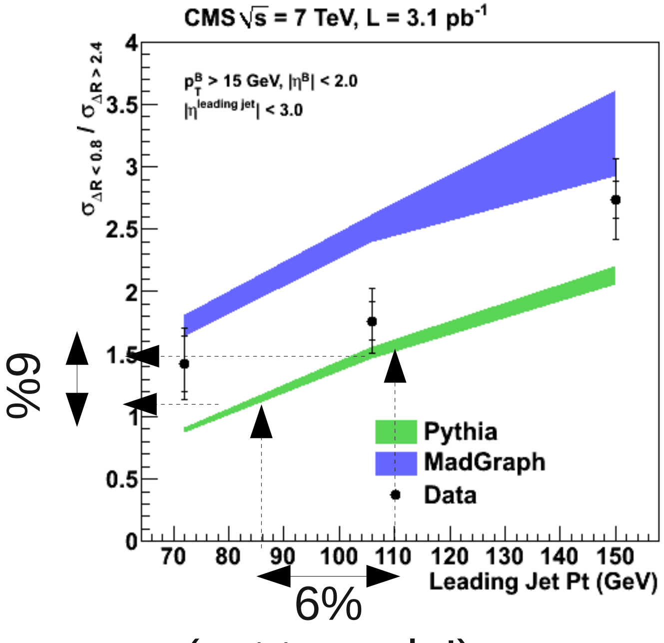
- The efficiency and purity correction are given for the (simulated) phase space defined as
  - Two B hadrons both with
    - pt > 15 GeV
    - |eta| < 2
- The reconstructed B are selected with
    - pT(sum of tracks@SV) > 8 GeV (about half of the simulated pt is reconstructed at the secondary vertex)
    - |eta| < 2
- The 8GeV cut is somehow arbitrary. The uncertainty arising from this choice of the cut is estimated by raising the cut to 10GeV, recomputing the efficiency correction and looking how the results would change.

| Lead.Jet | Pt > 56 | Pt > 84 | Pt > 120 |
|---|---|---|---|
| dR<0.8/ dR>2.4 | 0.97 | 0.97 | 0.99 |

# Jet Energy Scale

- JES uncertainty for our jet pt range (56 GeV to ~200 GeV) is at most 3%
  - 2.5% for most of the range
- The leading jet can be either a B or a light jet
  - Additional 5% uncertainty estimated using pythia vs Herwig  B-jet energy response
  - Also used in BPH-10-009 and TOP group
- GSP/FCR ratio increases with leading jet pT
  - JES uncertainty translates into a ratio uncertainty for a given LJ pt bin
- The result is a 6% systematic uncertainty on the ratio

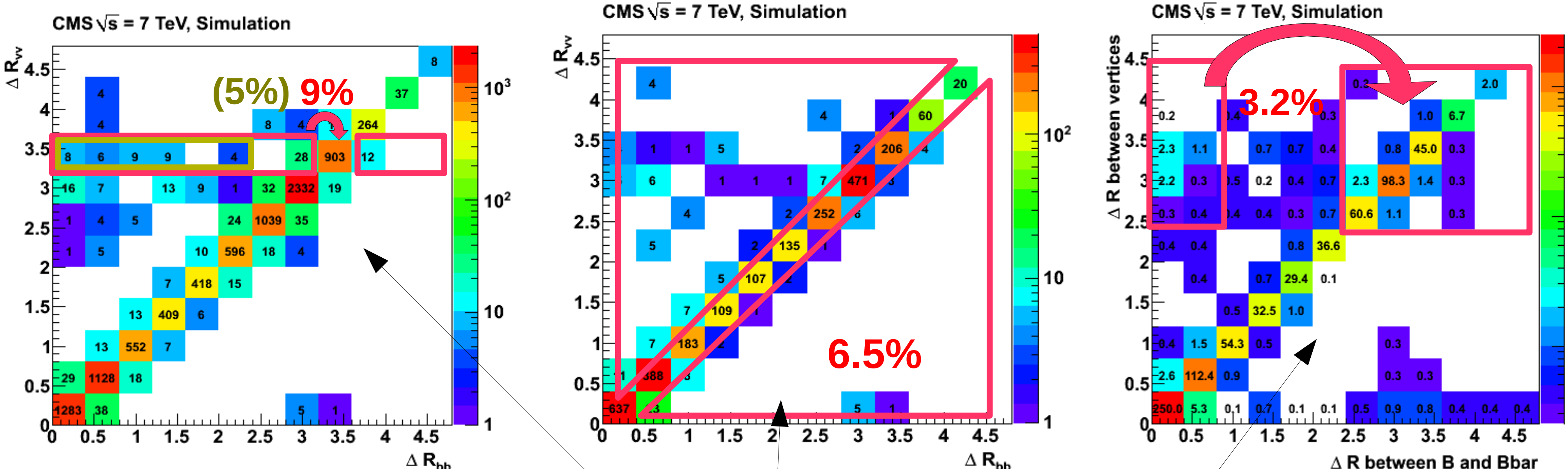



(not to scale!)

# Bin migration

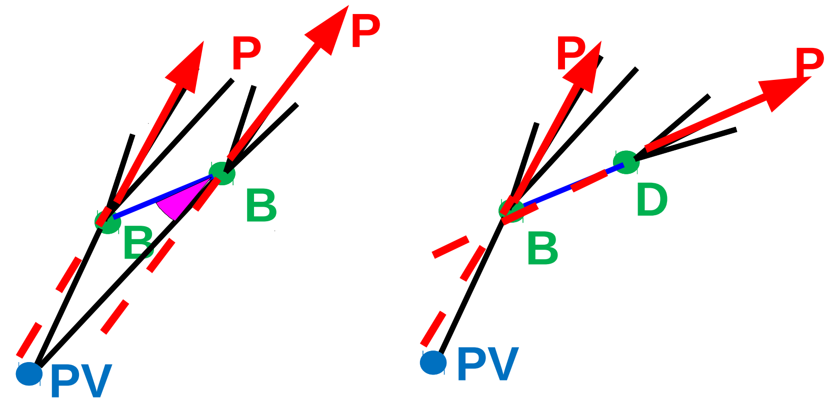
LJet $p_T$ > 56 | LJet $p_T$ > 84 | LJet $p_T$ > 120




- <u>Largest</u> migration into a single bin < 9% (5% excluding migration to next bins)
- Overall number of off-diagonal ~ 5%, 6.5%, 6.5% (Jet pt > 56, 84,120 GeV)
  - Excluding next-bin migration: 2%, 3%, 4.5%
- Migration from DR<0.8 to DR>2.4 region: 1.2%, 2.6%, 3.2%
  - Systematic uncertainty estimated by changing +-50% the GSP/FCR ratio
- Events with $|\Delta R_{vv}-\Delta R_{bb}|>0.4$ less than 2%
- Final effect on GSP/FCR ratio: **0.6%, 1.3%, 1.6%**

# B-Candidates selection

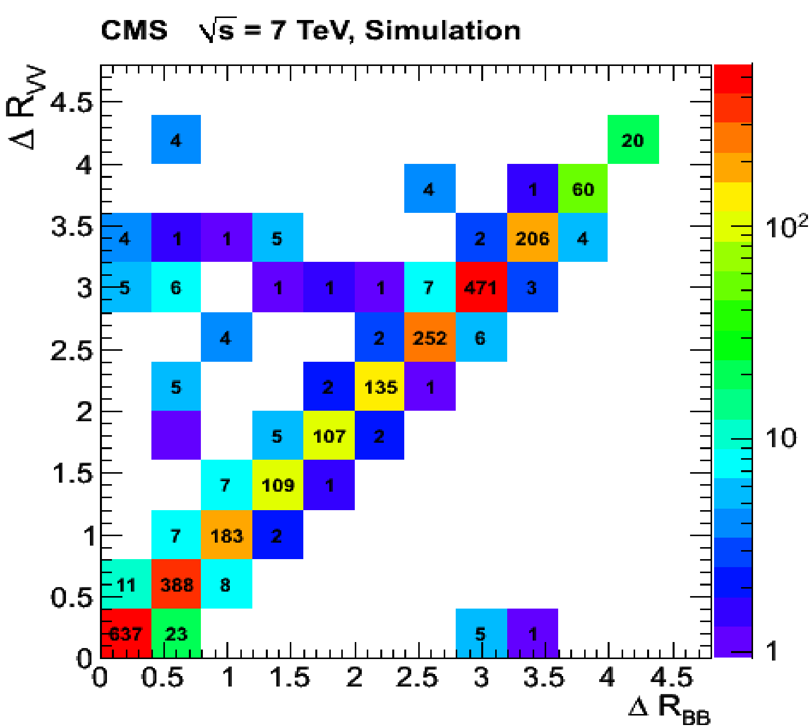
- Secondary Vertices from IVF can also originate from charm or light
- B decays can give more than one SV (B->D chain): possible background for small angle region
- Additional selection and merging of SV into ***B-candidate*** is done as follows:
  - SV are pre-selected if: m>1.4GeV, Ntracks >=3, dLenSig3D > 5, |eta| < 2, pt > 8 GeV
  - 2 SV are merged into a single “B-candidate” if: ΔR < 0.4, sumMass < 5.5, far-vertex (D) momentum pointing (cos > 0.99) to the near vertex(B)

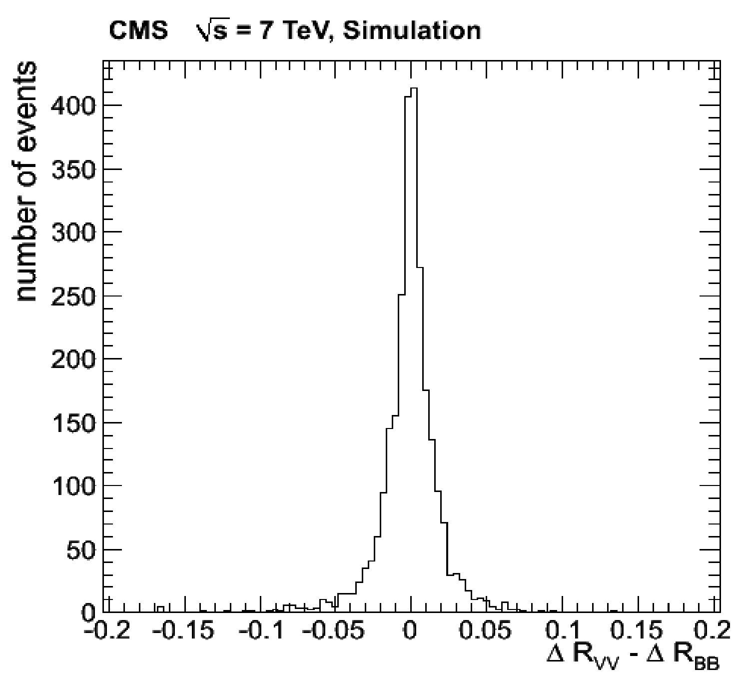



- Events with exactly two *B-candidates*, with sum of the masses > 4.5 GeV are kept for analysis

# Resolution and binning

- The angular resolution is evaluated on MC by looking at the reconstructed ΔR wrt the ΔR between the generator level B-hadrons
- A resolution of order 0.01-0.02 rad is observed
- The angular binning chosen for this analysis has a width of  0.4 that is much larger than the resolution. No unfolding needed to correct resolution effects
- Off diagonal contributions are mostly originating from events with one correctly reconstructed B and one fake, light or charm vertex.
  - These contributions are of order of 5% and are corrected bin by bin together with purity correction applied for non B background (see later slides for systematic uncert.)



# Efficiency

Efficiency corrections are derived from Monte Carlo simulation

The SV, and then B-candidate, reconstruction efficiency is determined by two effects, both can be parameterized as a function of the angular separation:

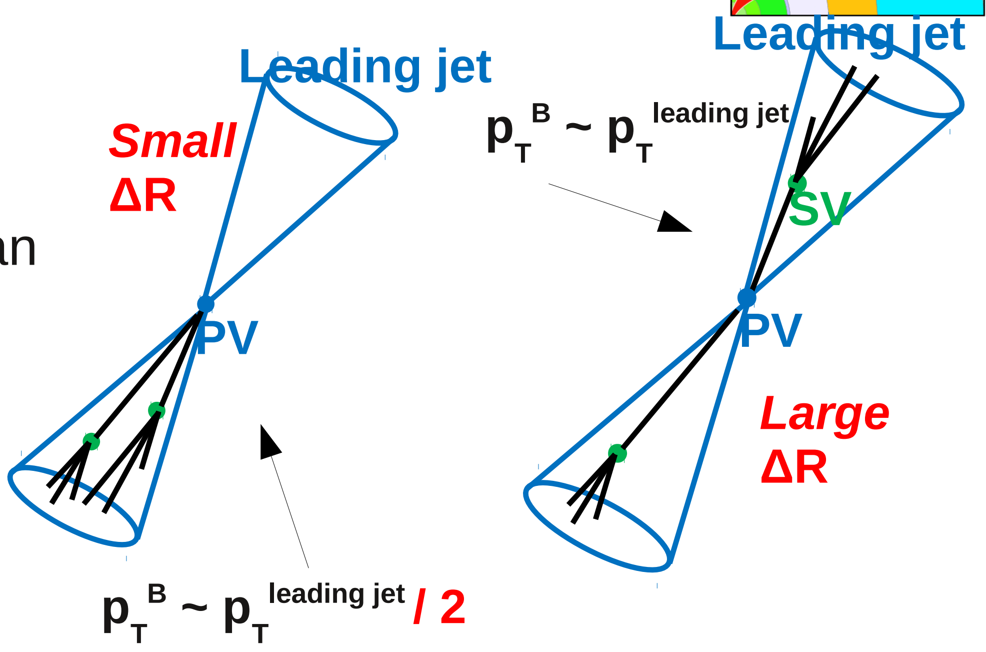
- pT of the B-hadrons (softer B are harder to reconstruct)

  At a given event energy scale (leading jet pt cut) the pT of the B-hadrons become a function of the opening angle (larger pT for back-to-back, smaller pT for collinear)

  - This effect is taken into account by computing the MC correction, as a function of ΔR, independently for each leading jet pt bin

- At small ΔR an algorithmic efficiency loss is expected for reconstructing two B hadrons

  The MC description of the algorithmic inefficiency is verified with a data mixing technique



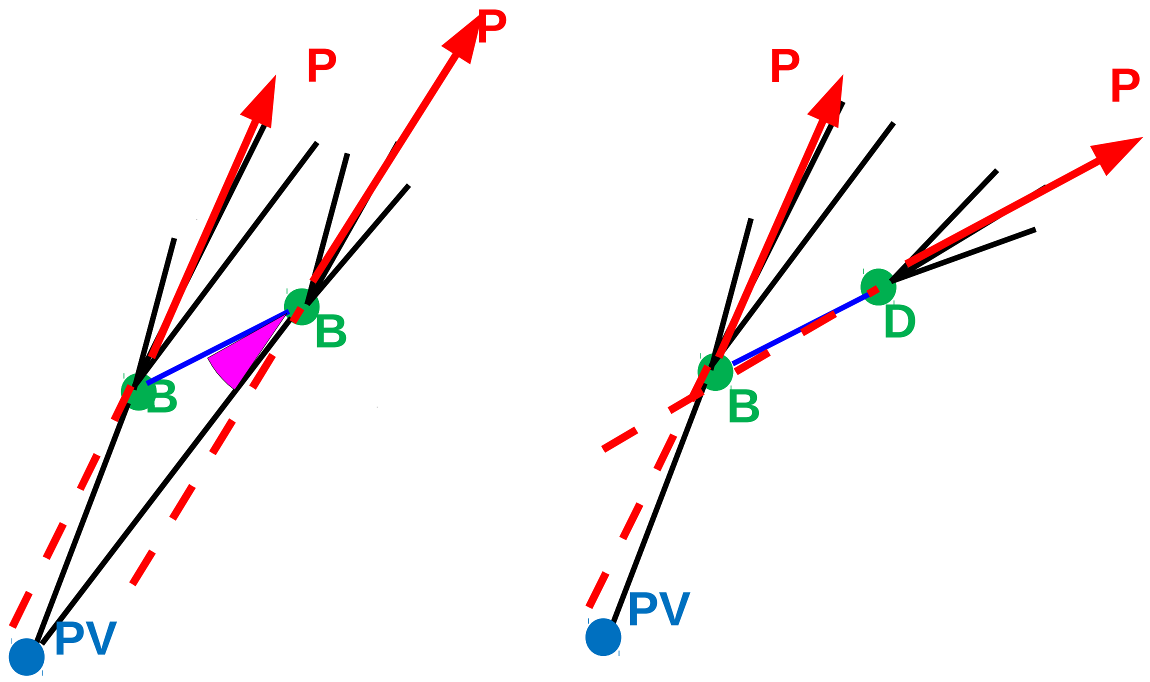
# Purity correction

- A purity correction is derived from MC and applied bin by bin, taking into account the following:
  - Bin migration
  - Background from events with 2 vertices from bc, bl, cc, cl
  - Background from events with more than 2 B
- A multiplicative factor correction is chosen (instead of number of events subtraction) because all contribution scale with Secondary Vertex reconstruction efficiency
- As the goal of this analysis is the measurement of the differential cross section (rather than the absolute value) it is important that the purity correction is reasonably flat
- From MC simulation we obtain purity within 80-90%
- The limited MC statistics introduce an uncertainty in the final purity/efficiency correction that can be mitigated with variable binning:
  - The largest between half bin-to-bin fluctuation and bin stat uncertainty is taken as systematic ( 13%)

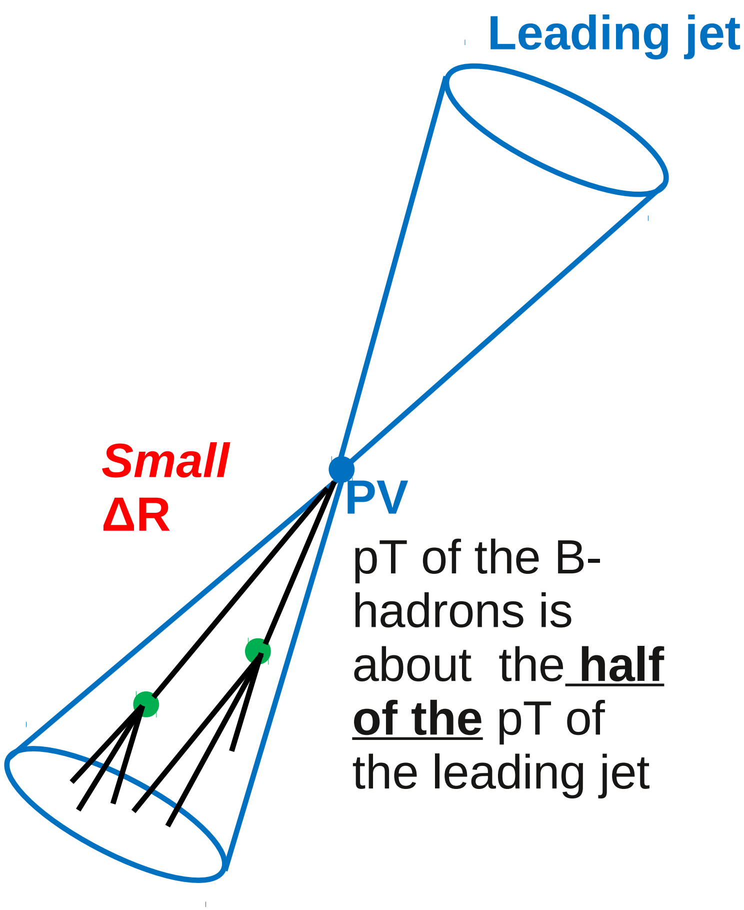
# Two B in one jet

- Standard b-tagging vertex finder (AVR based) uses the jet direction to preselect the tracks to fit
  - Does not work well if there are two B in one jet
- A **jet-independent** inclusive vertex finder (IVF) has been developed to reconstruct B-hadron secondary vertices without using the jet information
- The IVF algorithm works in few steps:
  - Seeding: select seed tracks with high I.P. wrt PV
  - Clustering: select tracks that are “compatible” with the seed
  - Fitting: use standard AVF/AVR to fit tracks into vertices
  - Merging: clean-up duplicate vertices (sharing tracks and being compatible in 3D position)
  - Arbitration: add tracks not previously selected if the compatibility to the SV is greater than that to the PV. Remove tracks that are more compatible with the PV than SV.

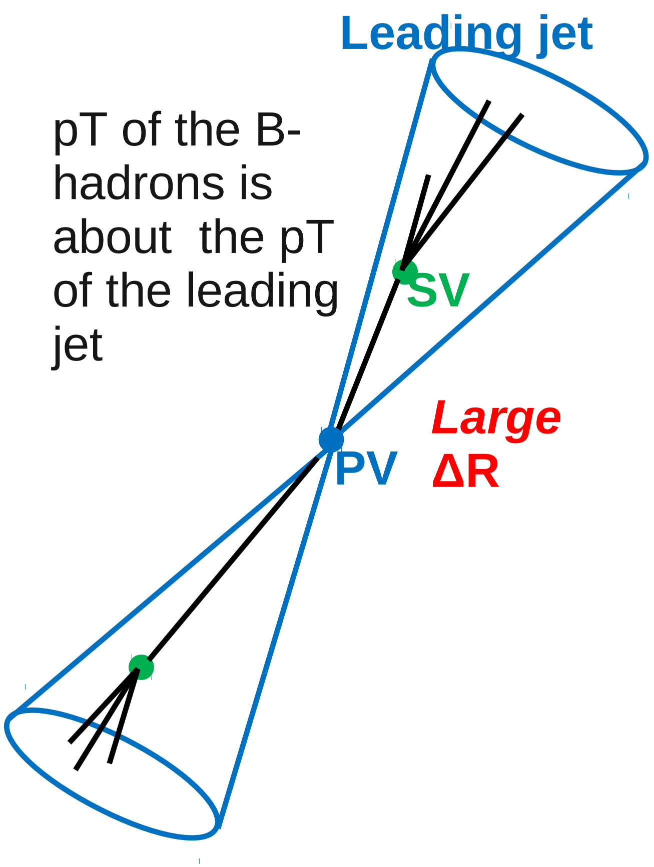
# B → D pointing angles



# pT vs ΔR trend at fixed leading jet pT




pT of the B-hadrons is about the **half of the** pT of the leading jet




pT of the B-hadrons is about the pT of the leading jet